

# Deep Learning for quantification and localization of multi-omics cell imaging datasets to interrogate mRNA and protein spatial and temporal subcellular interactions

BIOSS-IA 2020

**Emmanuel Bouilhol<sup>2,4</sup>**, Edgar Lefevre<sup>2</sup>, Anca F. Savulescu<sup>1</sup>, Robyn Brackin<sup>3</sup>, Benjamin Dartigues<sup>2</sup>, Hayssam Soueidan<sup>2</sup>, Macha Nikolski<sup>2,4</sup> and Musa M. Mhlanga<sup>1</sup>

<sup>1</sup>Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa.

<sup>2</sup>Université de Bordeaux, Bordeaux Bioinformatics Center, 33000, Bordeaux, France

<sup>3</sup>Advanced Medical Bioimaging, Charité – Universitätsmedizin, Berlin, Germany.

<sup>4</sup>Université de Bordeaux, CNRS, IBGC, UMR 5095, 33077, Bordeaux, France

# Context

Develop an analysis pipeline to understand how **mRNA** distribution is linked to **protein** distribution at the **single cell** scale

Data: cellular image data

→ Quantify and localize mRNAs and proteins

## Data Acquisition

Fluorescence Microscopy images  
of mRNA and Proteins



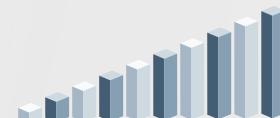
## Data Analysis

Image processing - Deep Learning



## Statistical Analysis

Develop statistical methods for  
spatial and temporal analysis





# Summary

- 1. Data Acquisition**
- 2. Data Analysis**
  - **Segmentation**
  - **Spot detection**
    - **Augmented microscopy**
    - **VAE-LSGAN architecture**
    - **Results**
- 3. Downstream Statistical Analysis**



# Summary

## 1. Data Acquisition

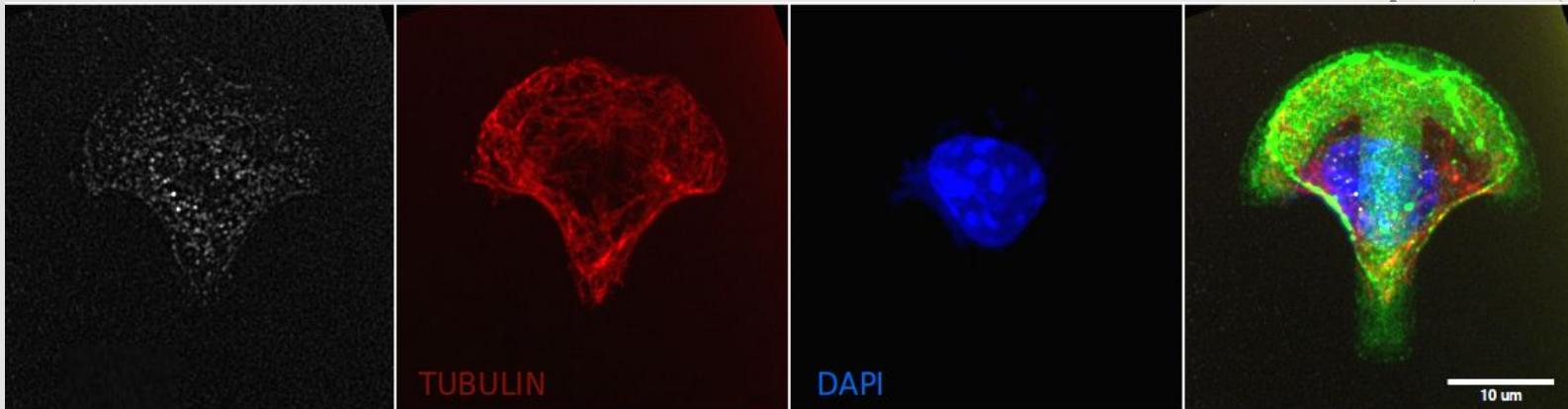
## 2. Data Analysis

- Segmentation
- Spot detection
  - Augmented microscopy
  - VAE-LSGAN architecture
  - Results

## 3. Downstream Statistical Analysis

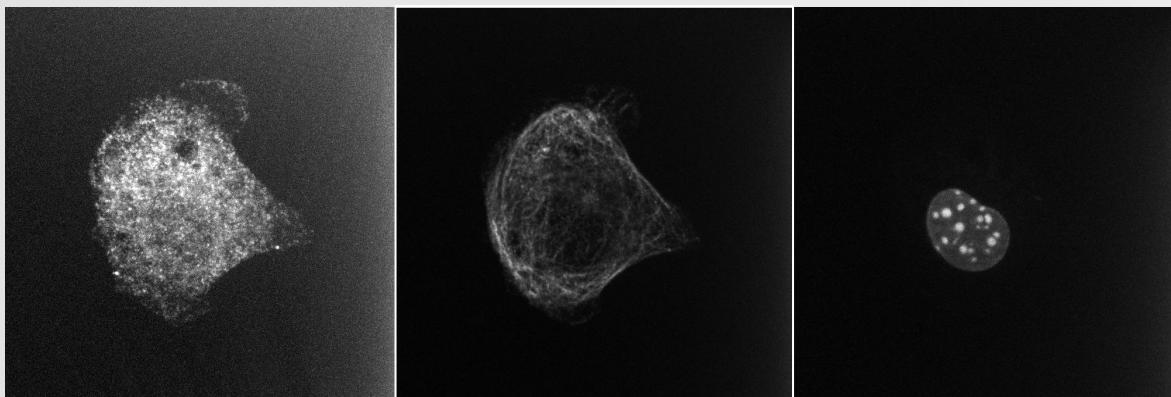
# Data Acquisition

FISH -> mRNA



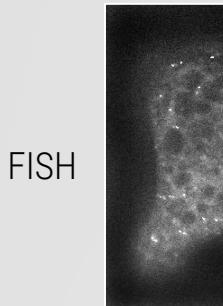
Micropattern

IF -> Proteins

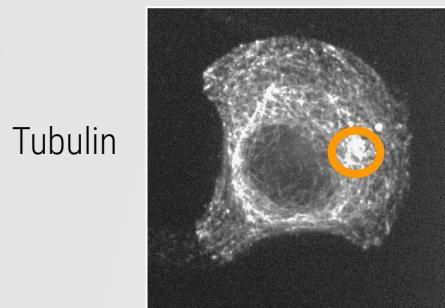
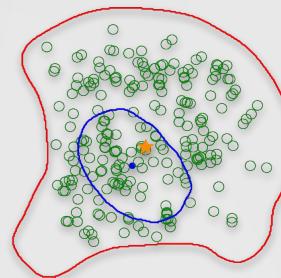


Collaboration with the  
MhlangaLab of the U. Cape  
Town - South Africa **1500**  
**micropatterned single cell**  
**images** (mouse fibroblasts)

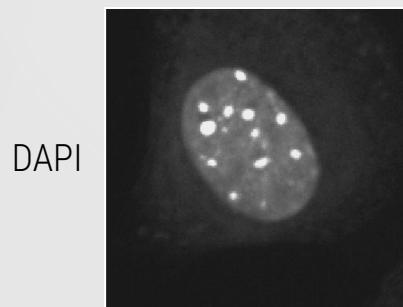
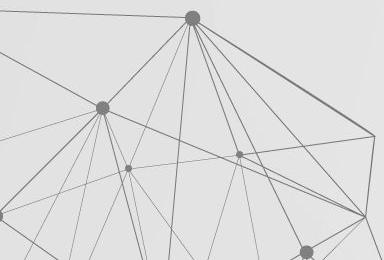
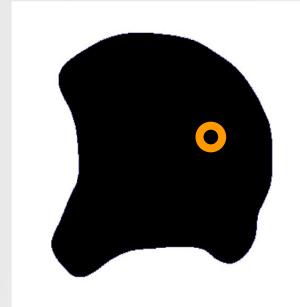
# Data Acquisition - features extraction



mRNA spots position (x,y)

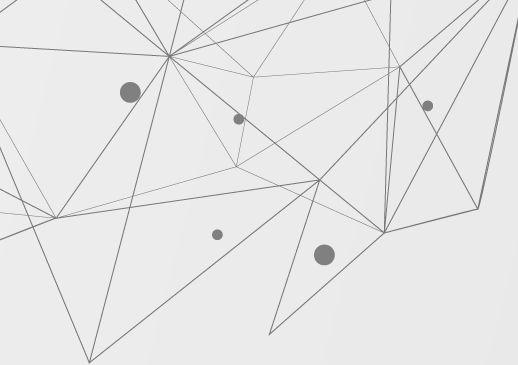


Cell Mask - MTOC (x,y)



Nucleus Mask





# Data Acquisition - difficulties

## Variation of SNR

---

Signal to Noise Ratio

## Artifacts, light aberrations

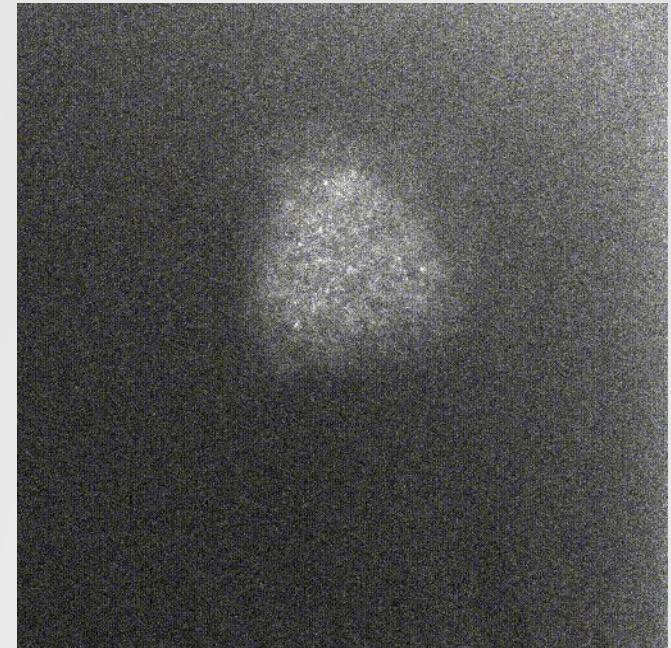
---

Presence of looking alike signals for spots or MTOCs

## Blur, out of focus

---

Difficulties to detect cell borders or slices out of focus





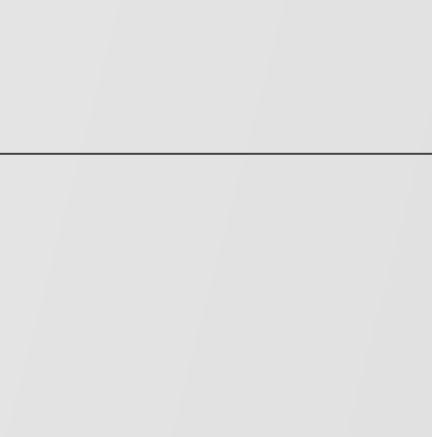
# Summary

1. Data Acquisition
  2. Data Analysis
  3. Downstream Statistical Analysis
- Segmentation
  - Spot detection
    - Augmented microscopy
    - VAE-LSGAN architecture
    - Results



---

# 1500 images



Classical algorithms of image processing:

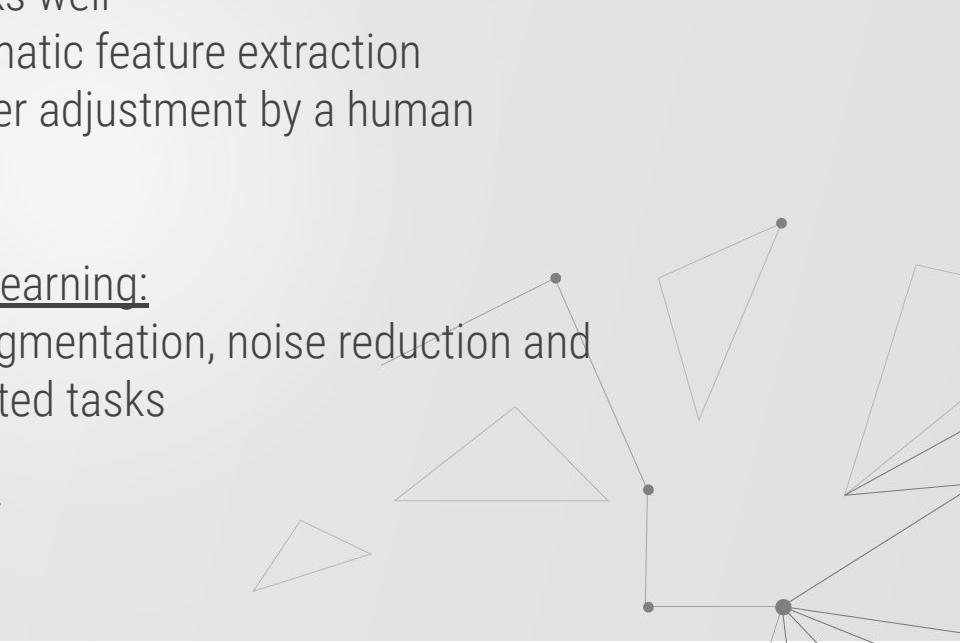
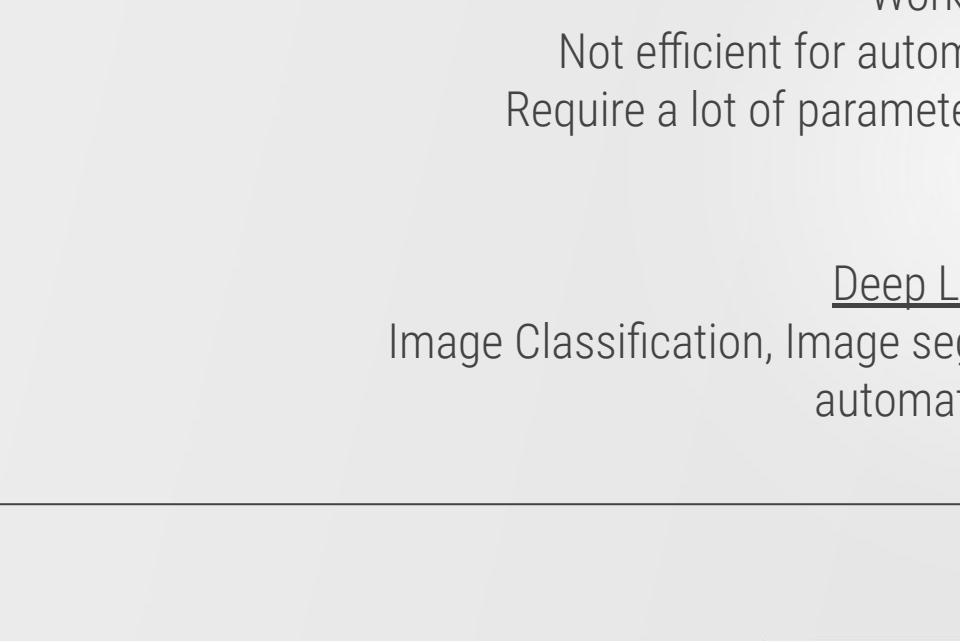
Works well

Not efficient for automatic feature extraction

Require a lot of parameter adjustment by a human

Deep Learning:

Image Classification, Image segmentation, noise reduction and  
automated tasks

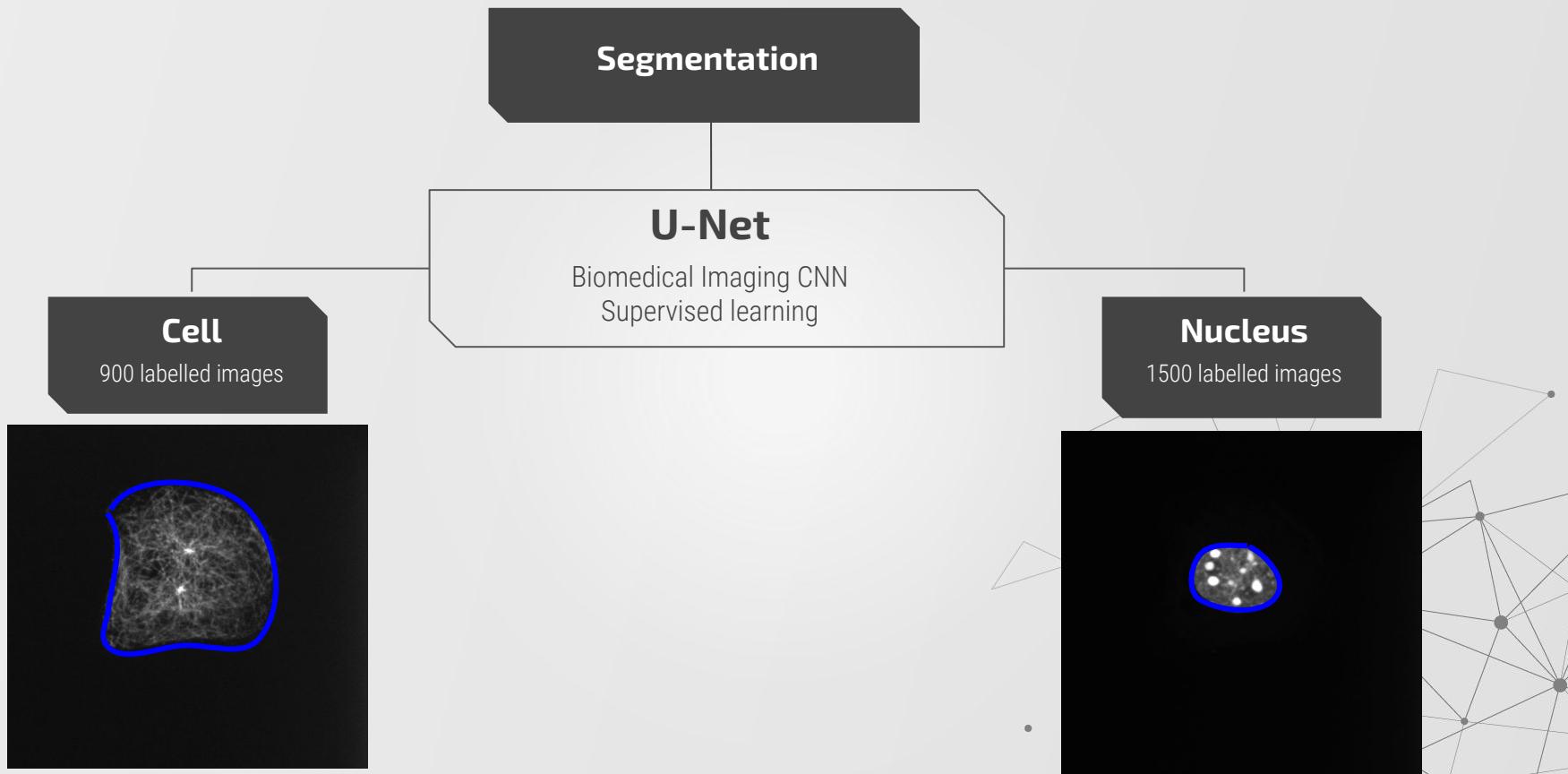




# Summary

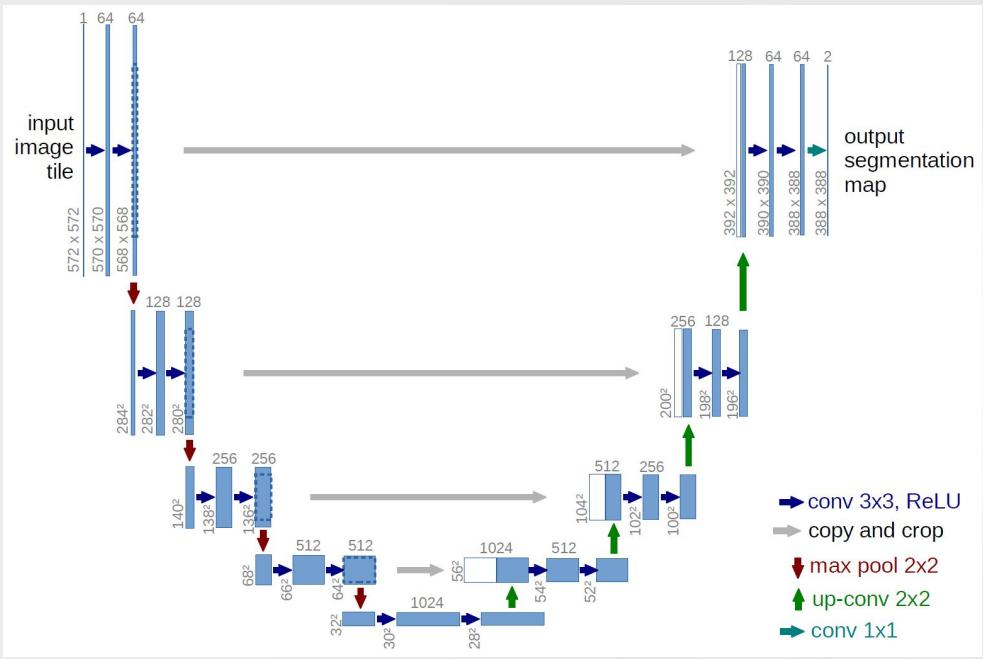
1. Data Acquisition
2. Data Analysis
  - Segmentation
  - Spot detection
    - Augmented microscopy
    - VAE-LSGAN architecture
    - Results
3. Downstream Statistical Analysis

# Data Analysis - Cell and nucleus segmentation



# Data Analysis - Cell and nucleus segmentation

## U-Net



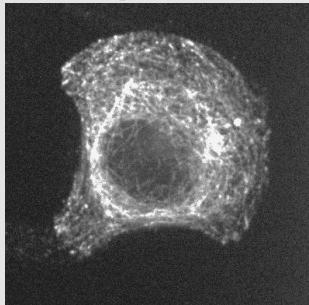
1. Can be trained with fewer images than many other networks

2. Capable of complete segmentation from fragmented annotation

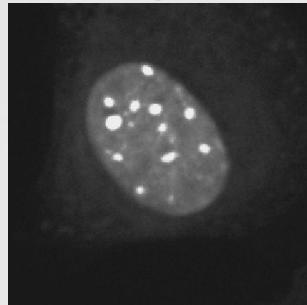
# Data Analysis - U-Net results

Images

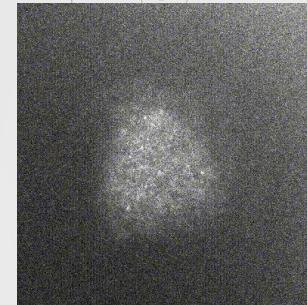
Cell segmentation



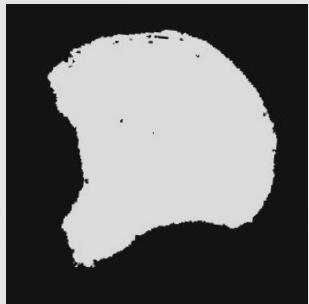
Nucleus segmentation



3D cell segmentation

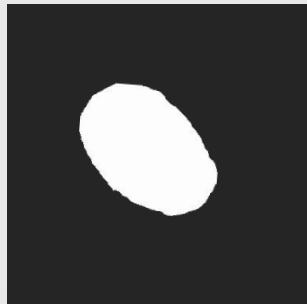


Mask  
Prediction

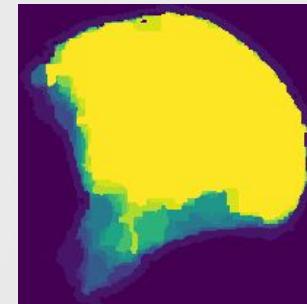


Dice  
Coefficient

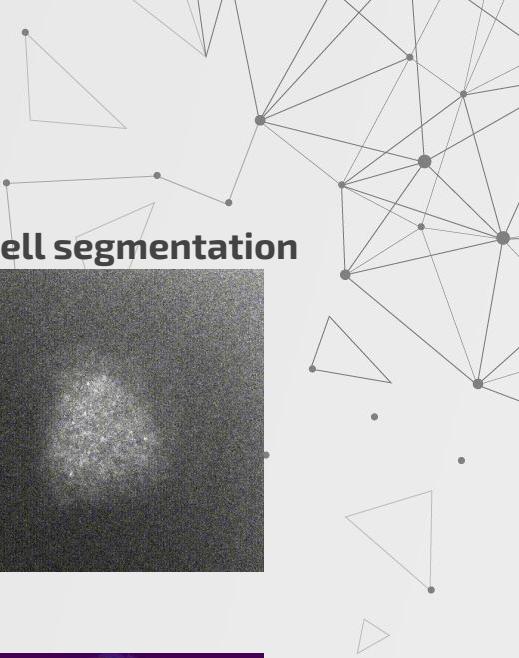
93 %



97 %



86 %

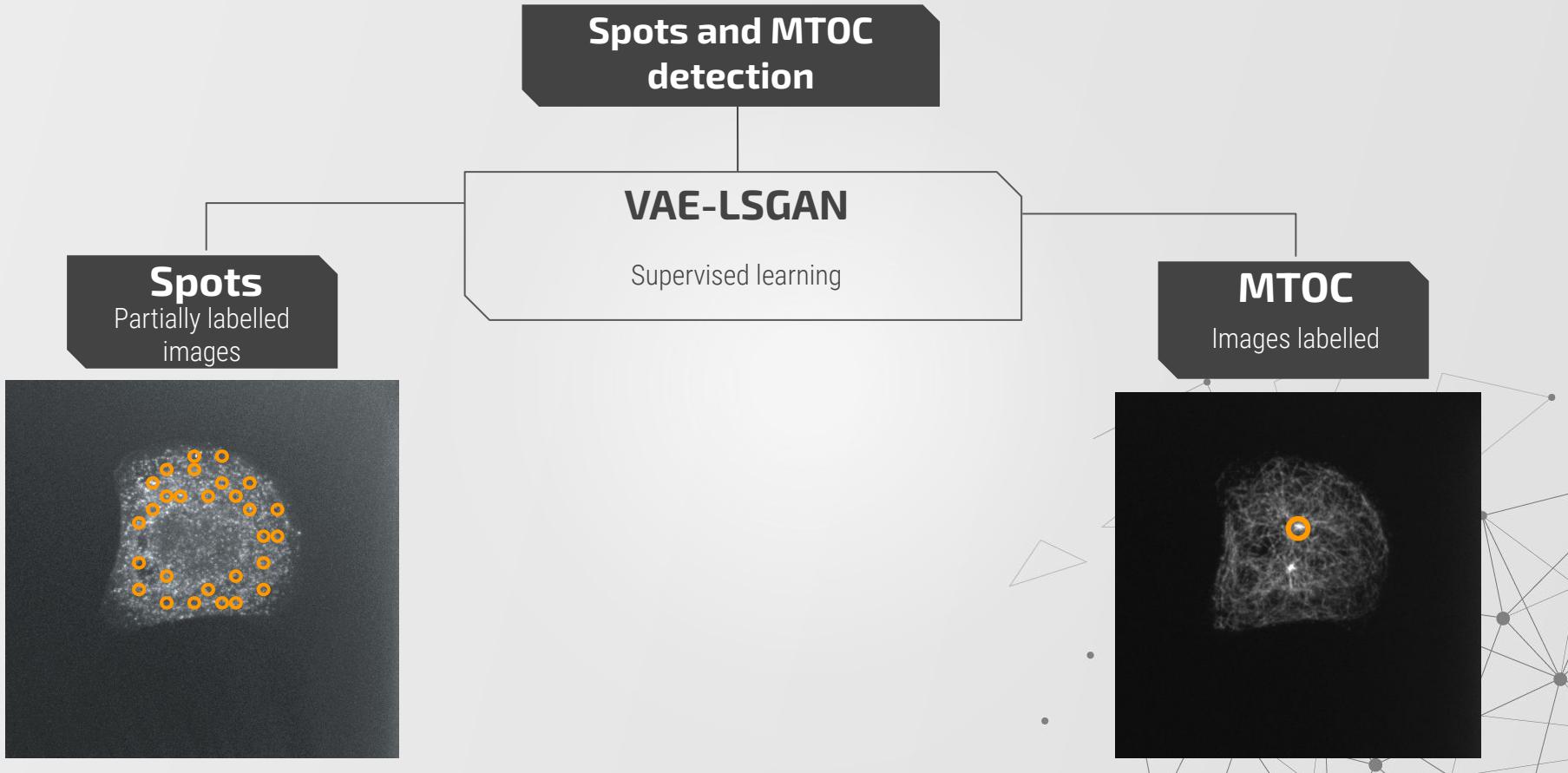




# Summary

1. Data Acquisition
2. Data Analysis
  - Segmentation
  - Spot detection
    - Augmented microscopy
    - VAE-LSGAN architecture
    - Results
3. Downstream Statistical Analysis

# Data Analysis - Deep Learning

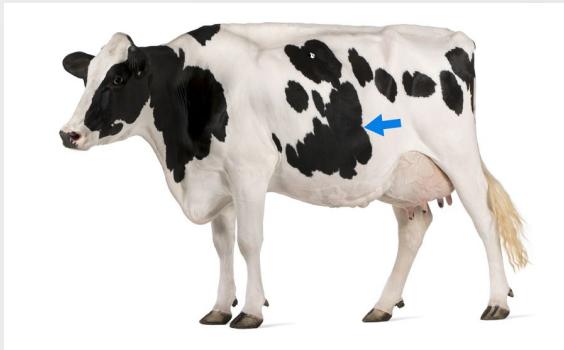


# Spot detection

## Definition of spot:

*"A small, usually round area of colour that is differently coloured or lighter or darker than the surface around it."*

- Dictionary of Cambridge

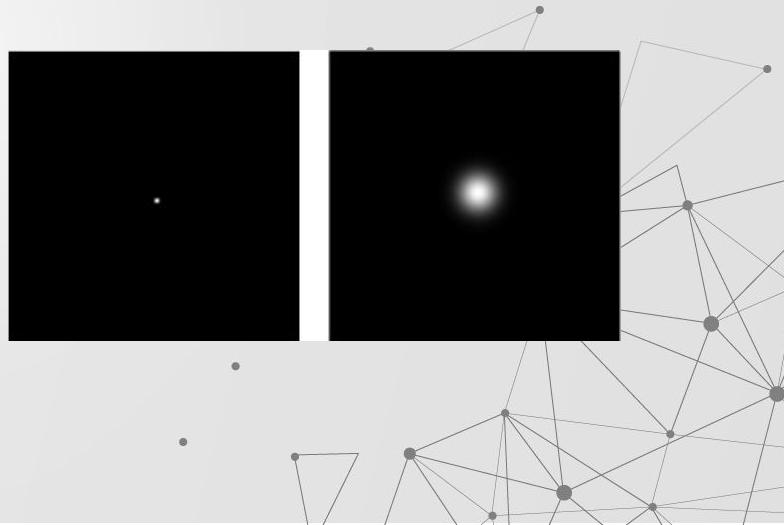


<https://dictionary.cambridge.org/fr/dictionnaire/anglais/spot>

## Definition of spot:

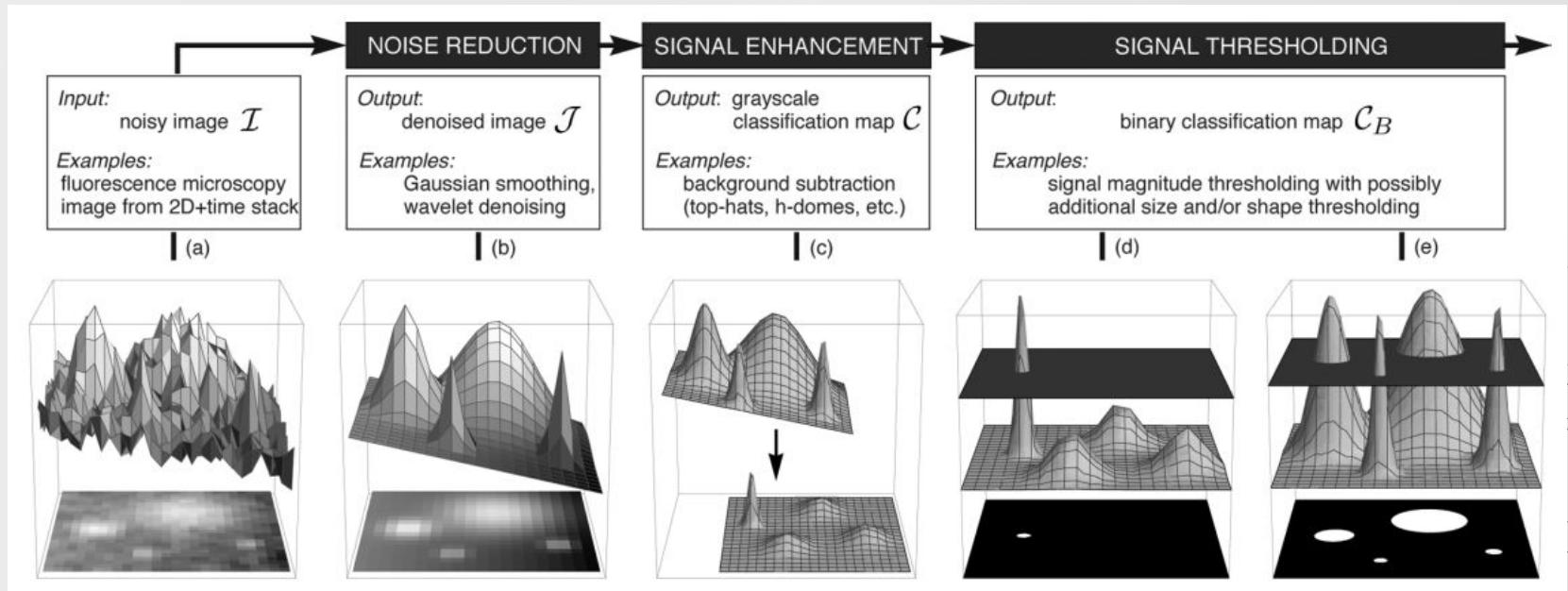
*"The specificity of mRNA spot is that they are necessarily small. mRNA spots should correspond to the Dirac delta function surrounded by the diffraction of the fluorescent signal."*

- Our definition



# Spot detection

Classical pipeline of spot detection



Smal, Ihor, et al. "Quantitative comparison of spot detection methods in fluorescence microscopy." IEEE transactions on medical imaging 29.2 (2009): 282-301.

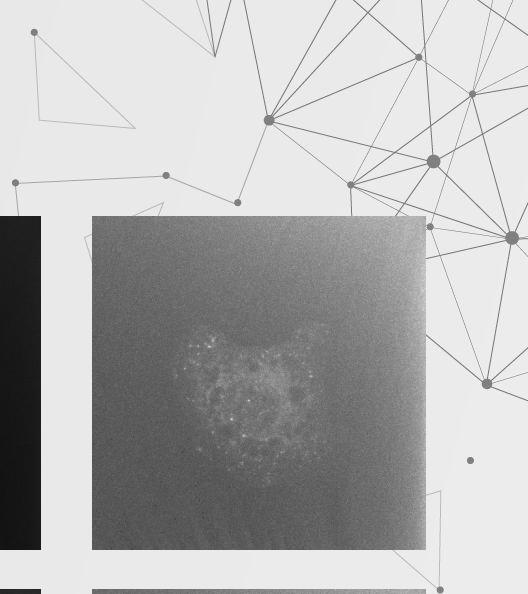
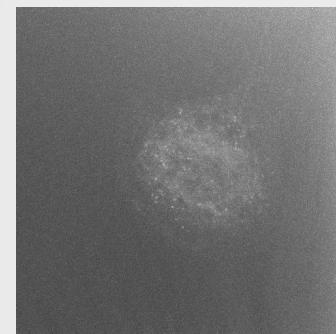
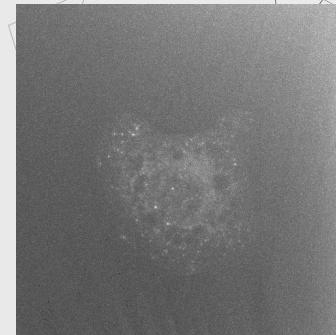
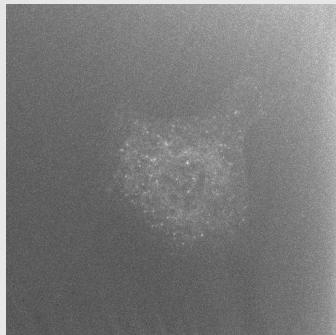
# Data Analysis - Challenge

Automatic detection is challenging:

Difficult to **label enough data** to train fully supervised learning algorithms

U-Net or similar classical Deep Learning approaches are not applicable here

Look into generative network



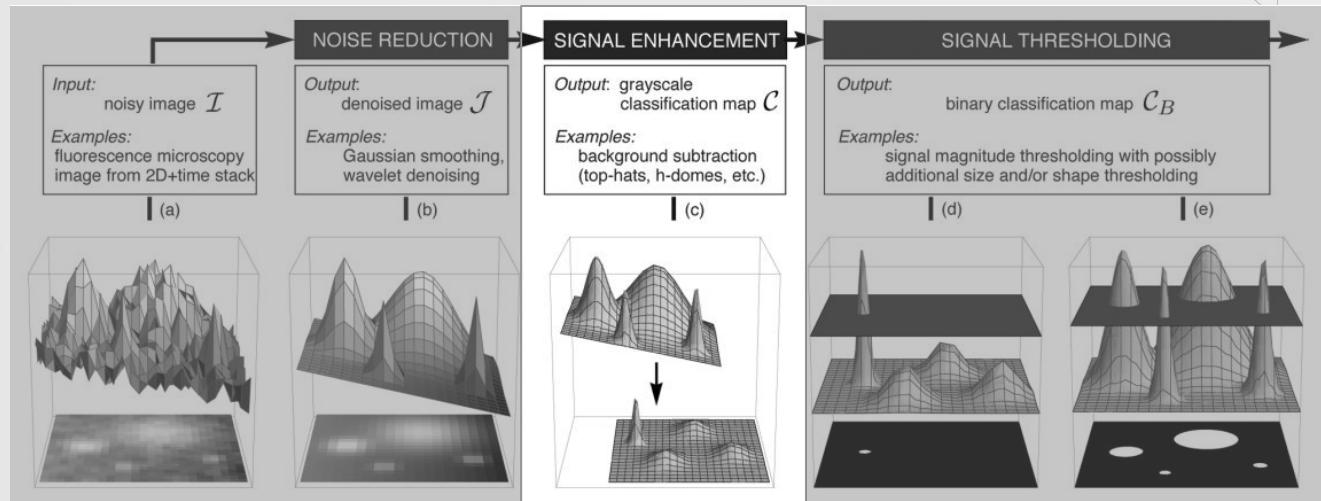
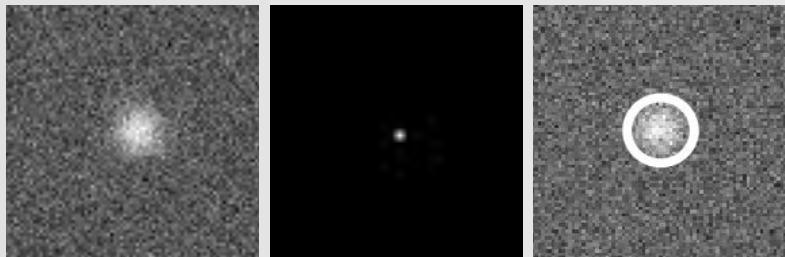


# Summary

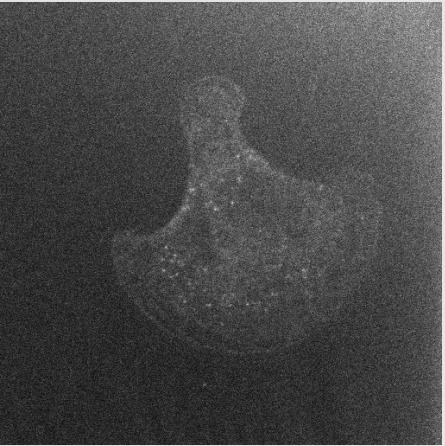
1. Data Acquisition
2. Data Analysis
  - Segmentation
  - Spot detection
  - Augmented microscopy
  - VAE-LSGAN architecture
  - Results
3. Downstream Statistical Analysis

# Augmented microscopy

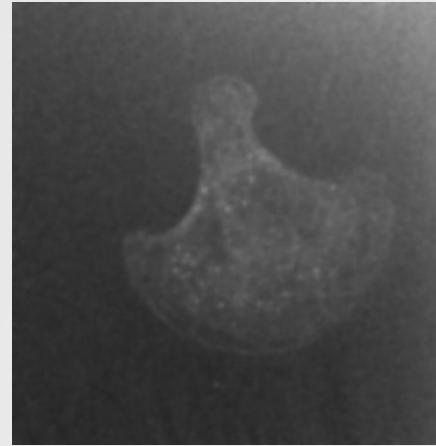
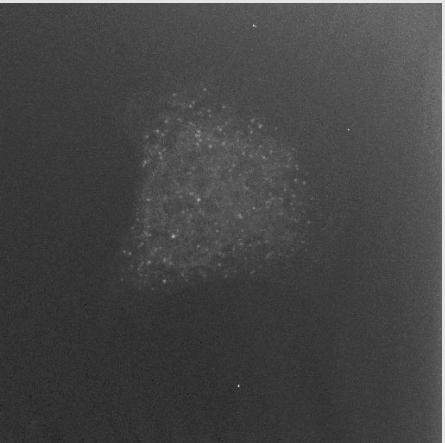
Signal enhancement:  
Enhance the region of interest



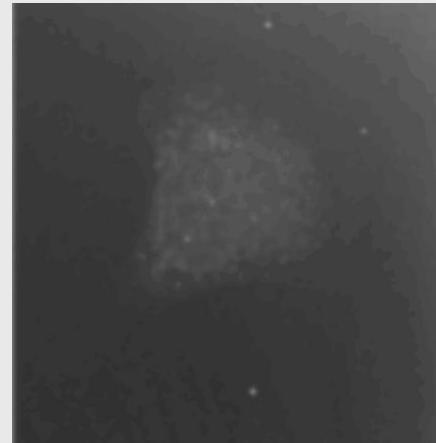
# Data Analysis - Challenge



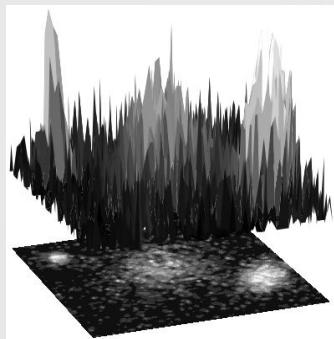
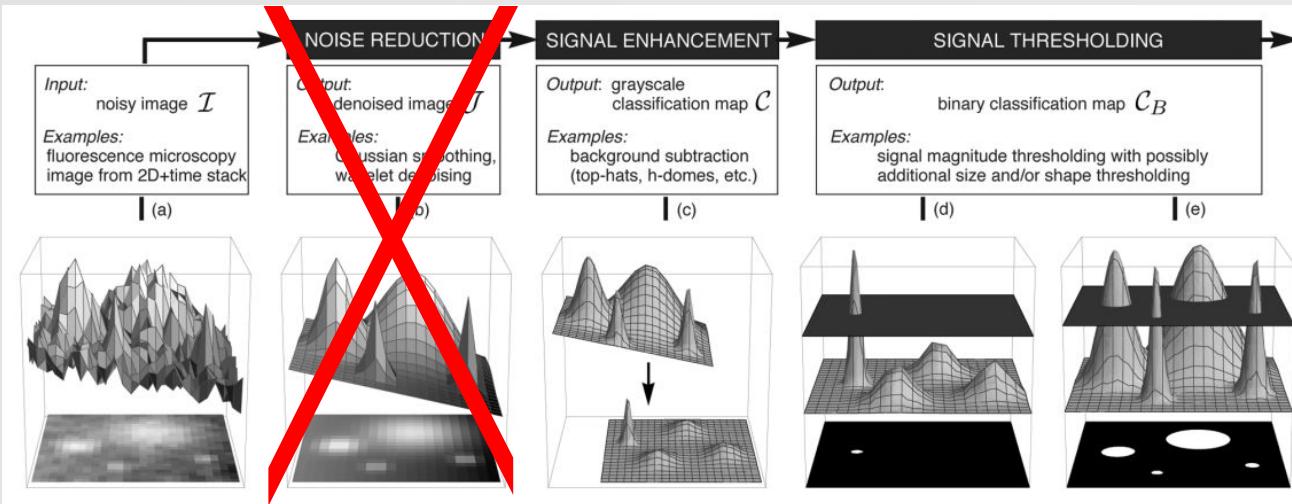
Noise reduction / Signal loss



Gaussian filter

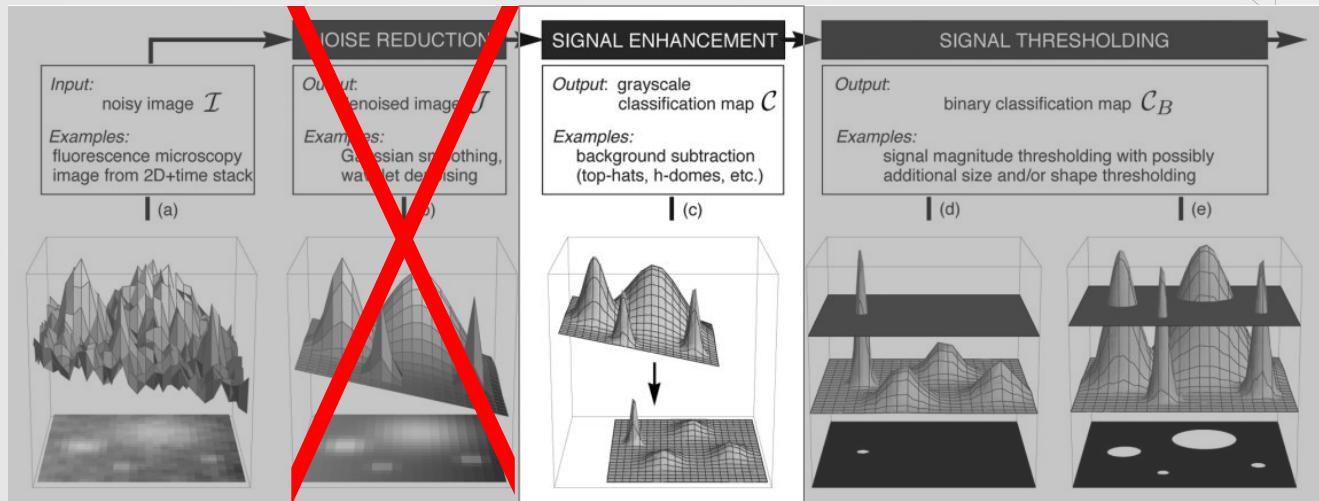
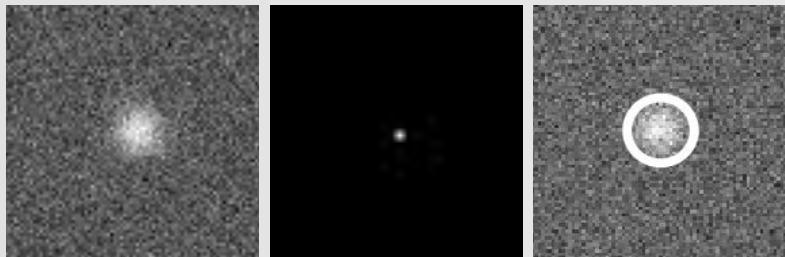


# Data Analysis - Challenge



# Augmented microscopy

Signal enhancement:  
Enhance the region of interest





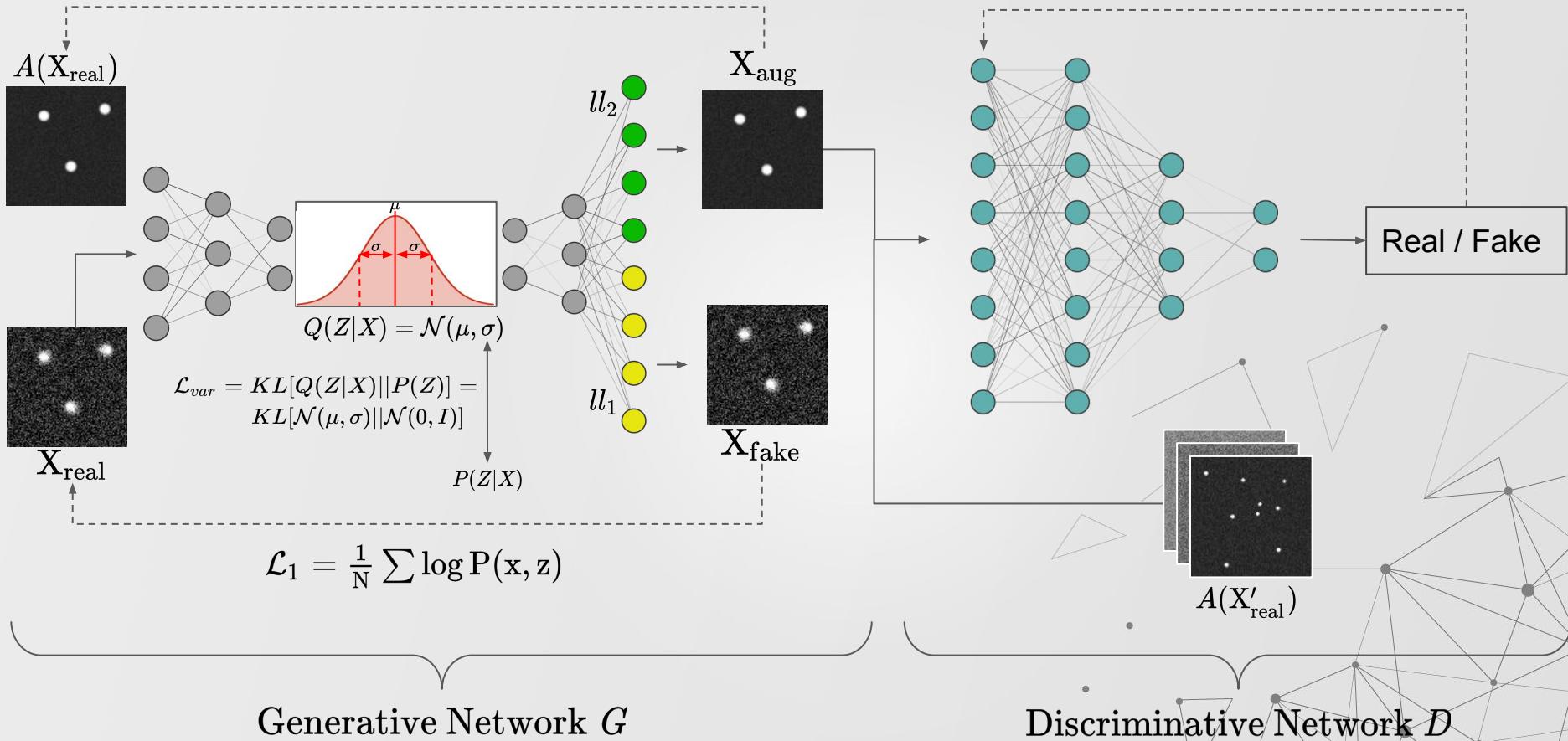
# Summary

1. Data Acquisition
2. Data Analysis
  - Segmentation
  - Spot detection
    - Augmented microscopy
    - VAE-LSGAN architecture
    - Results
3. Downstream Statistical Analysis

# VAE-LSGAN

$$\mathcal{L}_2 = \frac{2}{N} \sum \log P(x, z)$$

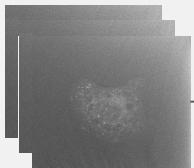
$$\mathcal{L}_D = D(x|z) + (1 - D(G(x))$$



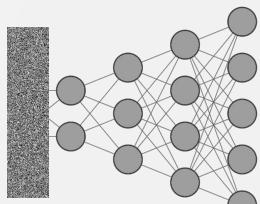
# Data Analysis - Our solution

**GAN**

Sample of real images  
(Target distribution)

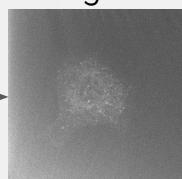


Input  
random  
noise  
vector

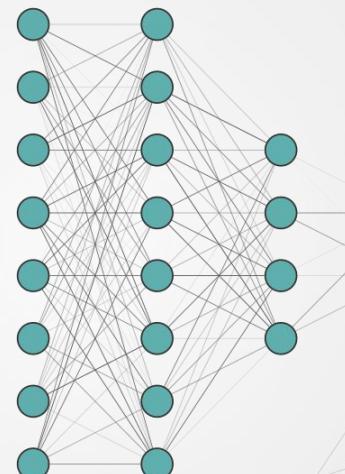


Generative  
network (G)

Generated  
images



Discriminative network (D)



Classification

Fake

$y = 0$

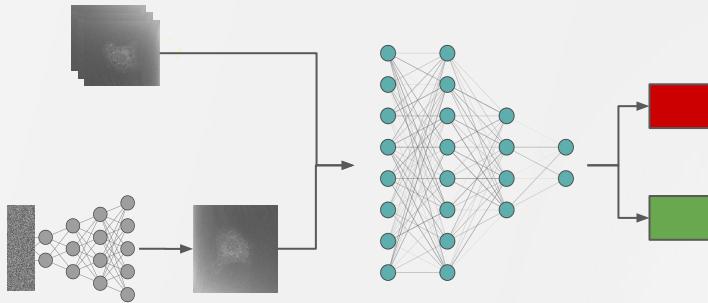
Real

$y = 1$

D learns the boundary decision between classes (fake / real)  
G learns to produce data similar to the input to fool D

# Data Analysis - Our solution

## GAN



The goal is to reach an equilibrium where D outputs 0.5 for each class (indecisive)

Difficult to train :

- Non-convergence
- Mode collapse
- Vanishing gradient

MiniMax game between the Generator (G) and the Discriminant (D)

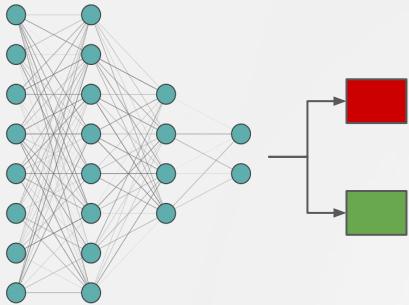
Objective function:

- Maximize the log of predicted probability of real images
- Maximize the log of the inverted probability of fake images

$$L = \min_G \max_D [\log(D(x)) + \log(1 - D(G(z)))]$$

# Data Analysis - Our solution

LSGAN

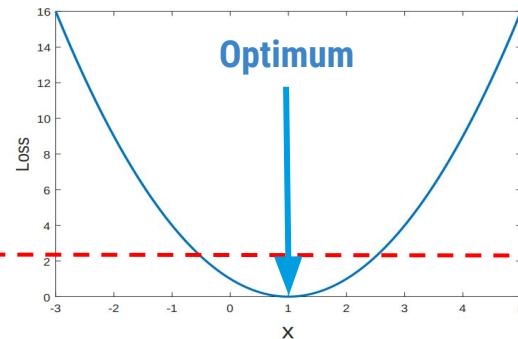
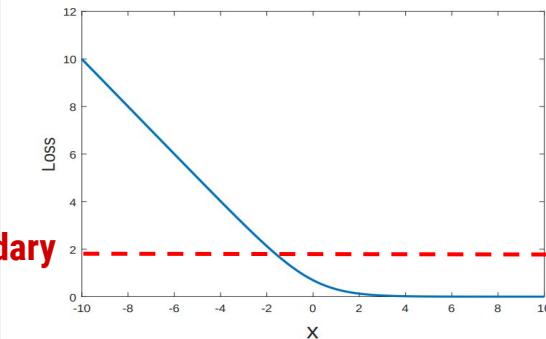


Discriminator loss:  
Binary Cross Entropy  $\rightarrow$  Least Square

Penalize fake samples even on the correct side of decision boundary: **generate data closer to real data**

Improved training stability

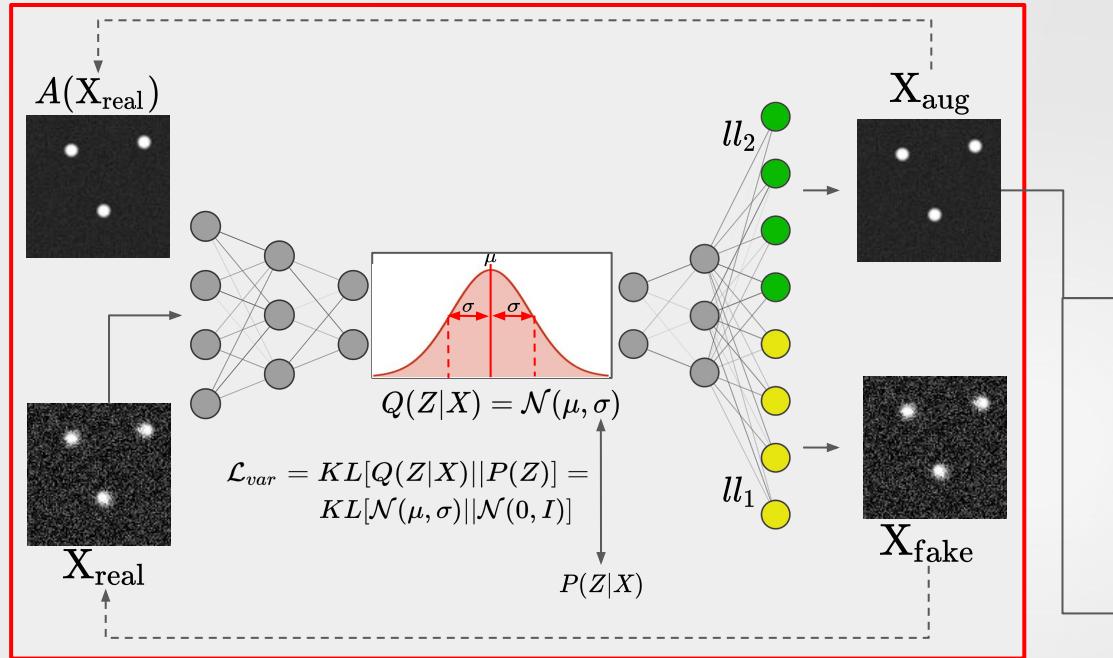
Decision boundary



# VAE-LSGAN

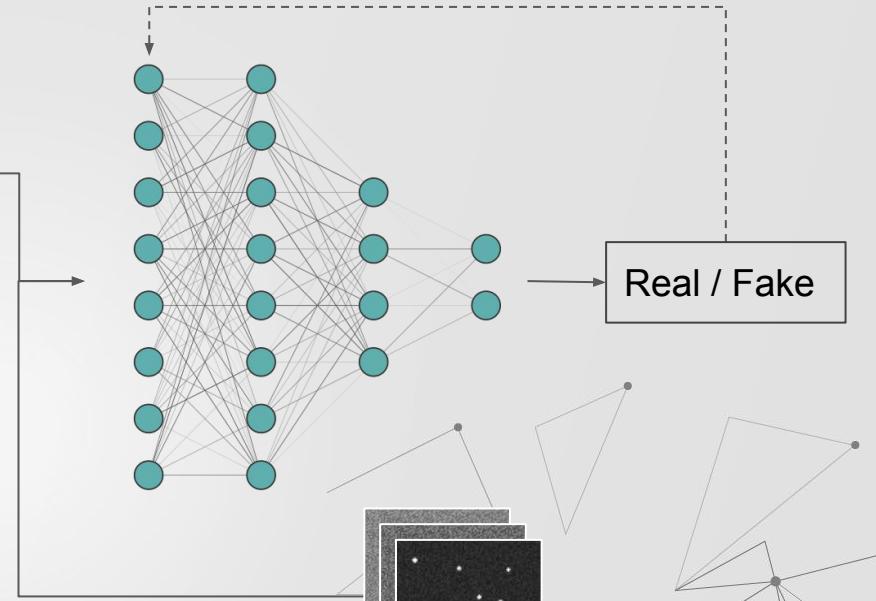
$$\mathcal{L}_2 = \frac{2}{N} \sum \log P(x, z)$$

$$\mathcal{L}_D = D(x|z) + (1 - D(G(x))$$



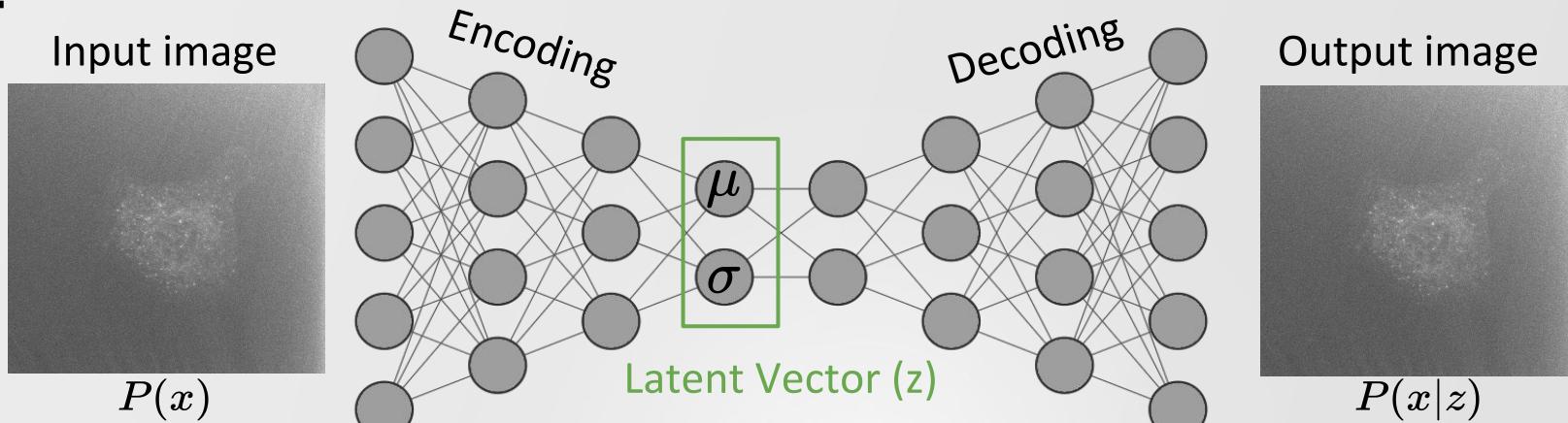
$$\mathcal{L}_1 = \frac{1}{N} \sum \log P(x, z)$$

Generative Network  $G$



Discriminative Network  $D$

# VAE



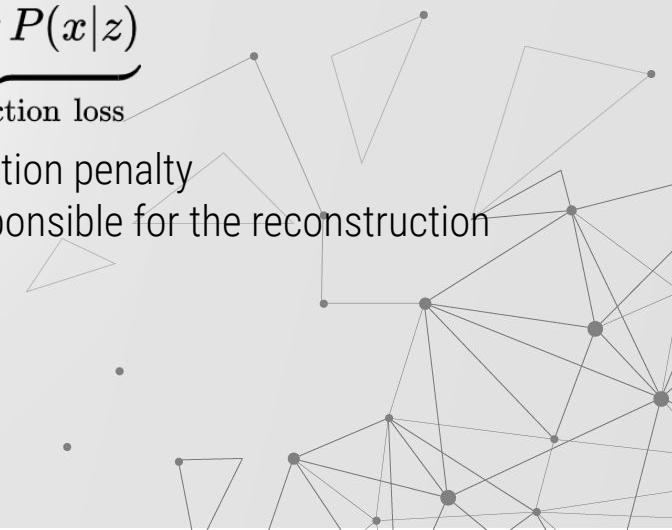
$$\mathcal{L} = \underbrace{\text{KL}[N(\mu, \sigma) \parallel N(0, I)]}_{\text{variational loss}} + \underbrace{\frac{1}{n} \sum \log P(x|z)}_{\text{reconstruction loss}}$$

VAE loss:

- KL = Kulback-Leibler Divergence, responsible for the regularization penalty
- Expected log-likelihood between input and output images, responsible for the reconstruction penalty

VAE learns:

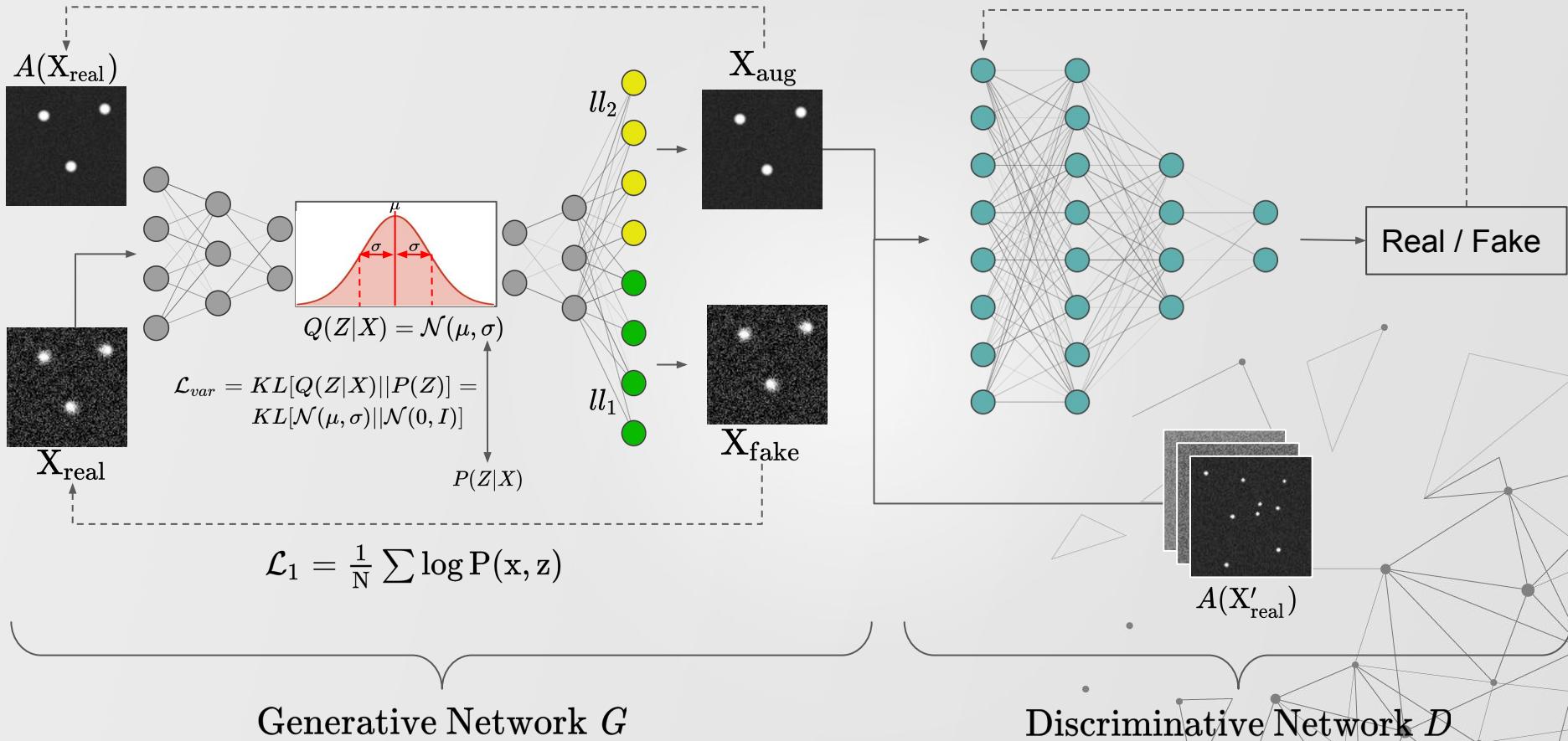
- A multi-modal **distribution that represents the data**
- To **reconstruct the input data**



# VAE-LSGAN

$$\mathcal{L}_2 = \frac{2}{N} \sum \log P(x, z)$$

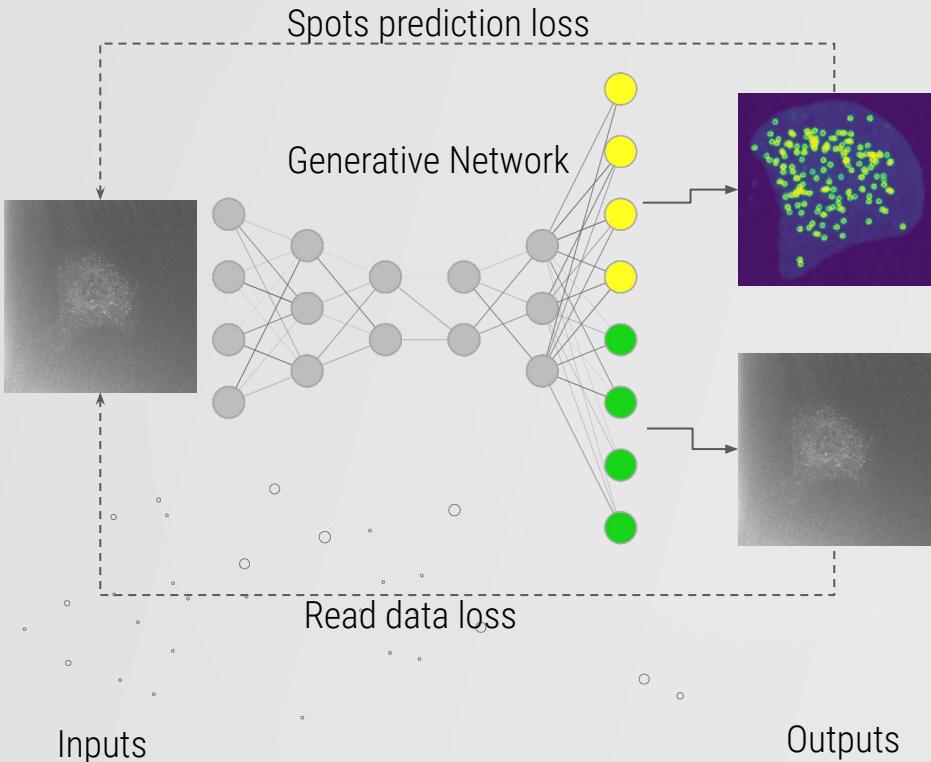
$$\mathcal{L}_D = D(x|z) + (1 - D(G(x))$$



# Data Analysis - Our solution

## VAE-LSGAN

### Training



We trained **two different last layers**:

- one to output similar image to real data,
- one to add yellow circles around spots

The objective function of this network is :

$$\text{VAE loss} = \text{real data loss} + \text{spot prediction loss} * 2$$

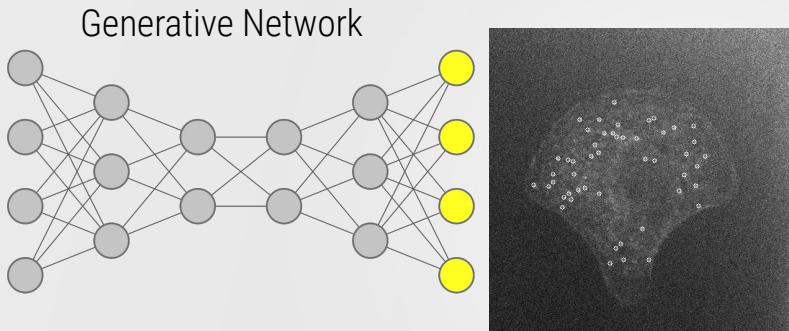
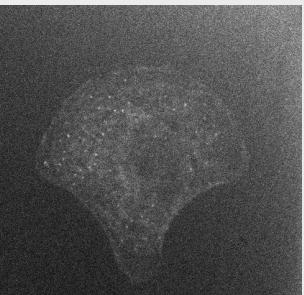
By penalizing the spot prediction loss, we force the network to add circles around spots when passing through the “green” layer



# Data Analysis - Our solution

## VAE-LSGAN

### Prediction



For prediction we only keep the VAE with the spot prediction layer

Detect spots with a classic blob detection (Determinant of Hessian)



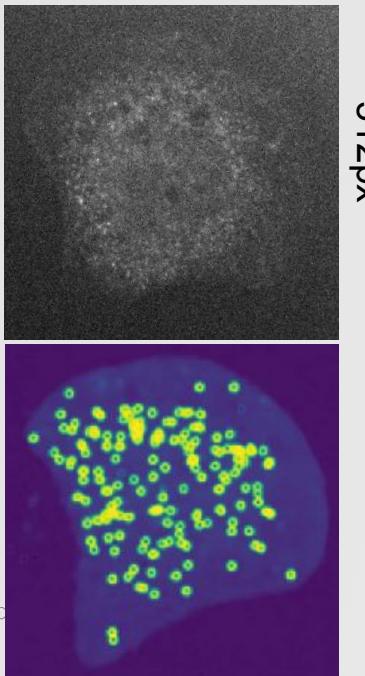


# Summary

1. Data Acquisition
2. Data Analysis
  - Segmentation
  - Spot detection
    - Augmented microscopy
    - VAE-LSGAN architecture
  - Results
3. Downstream Statistical Analysis

# VAE-LSGAN results

mRNA detection



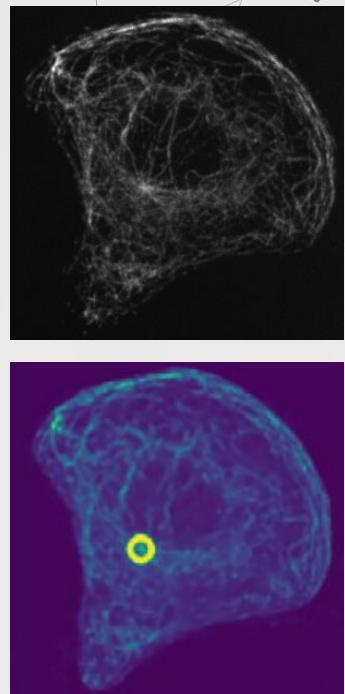
Images

Prediction

Test performances  
(spot density over cell)

95 %

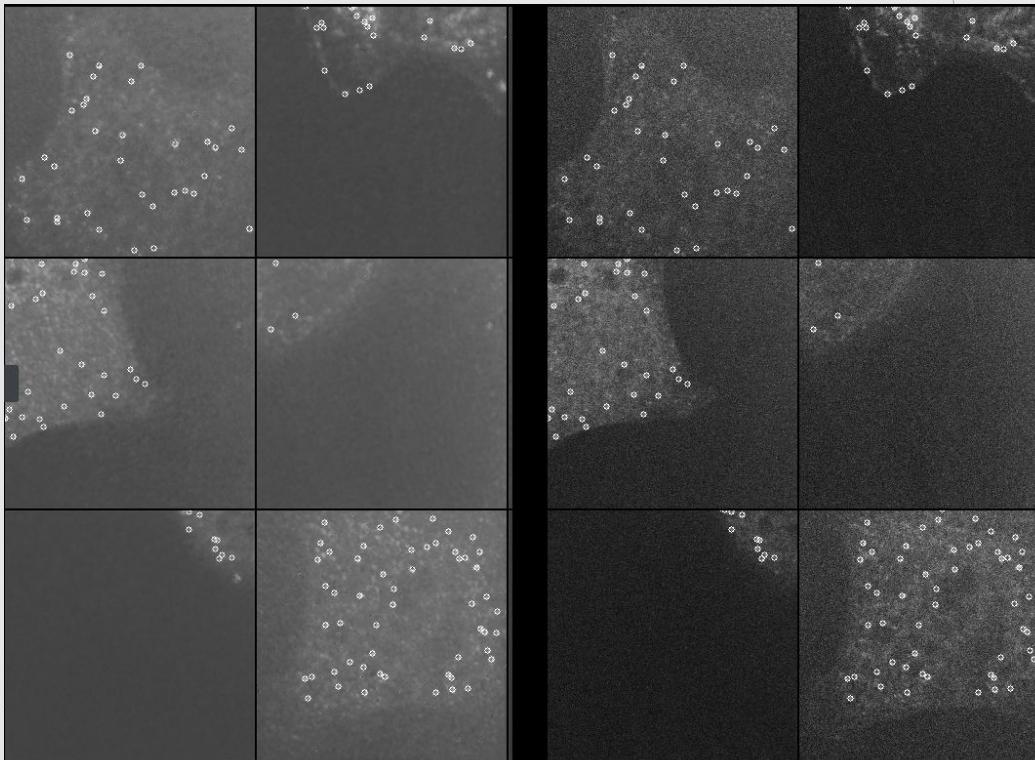
MTOC Detection



9.63

Mean Euclidean distance  
from Ground Truth

# VAE-LSGAN results

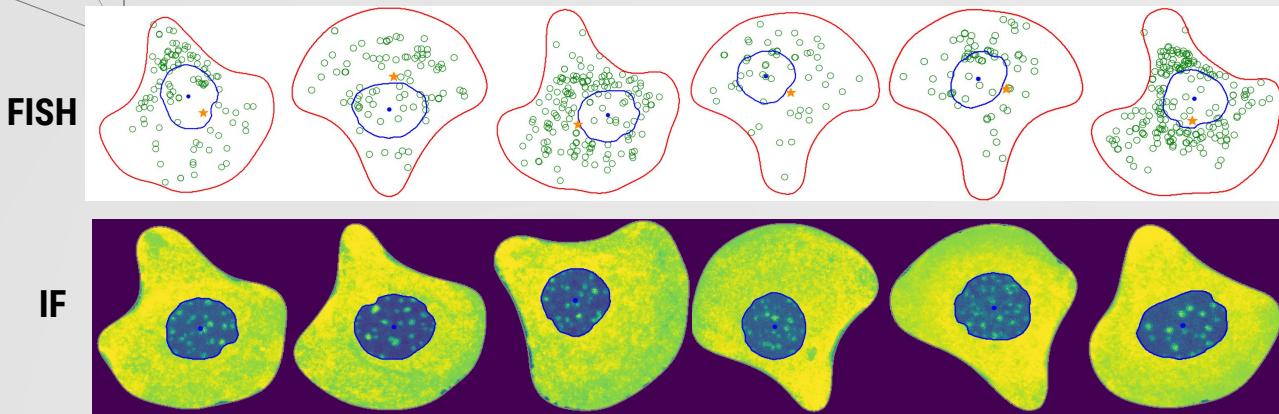




# Summary

1. Data Acquisition
2. Data Analysis
  - Segmentation
  - Spot detection
    - Augmented microscopy
    - VAE-LSGAN architecture
    - Results
3. Downstream Statistical Analysis

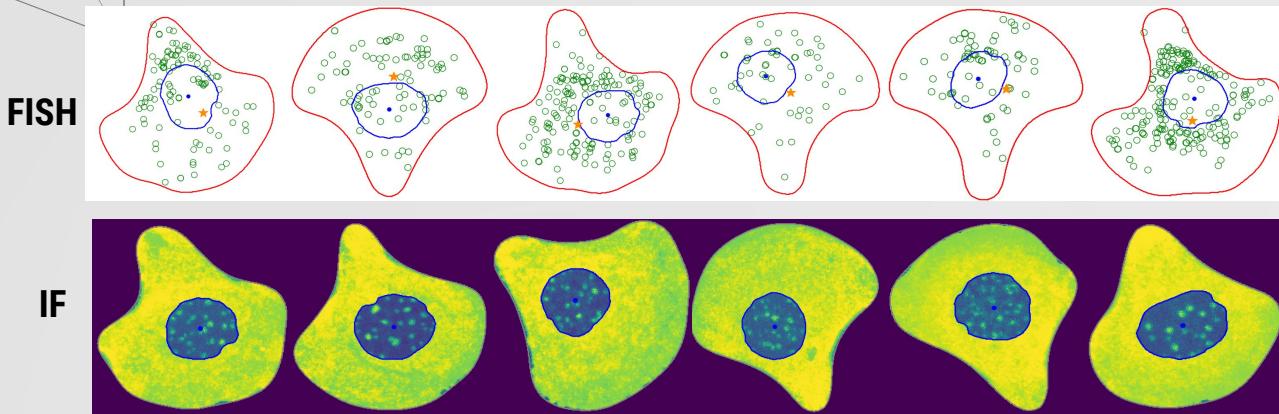
# Statistical analysis



Series of images with mRNA position and protein signal intensity, cell and nucleus area, we can compute some statistics:

- Quantitative information on mRNAs and proteins
- Diffusion in the cytoplasm
- Diffusion around the nucleus - at the cell periphery
- Diffusion according to MTOC's position
- Clustering mRNA / Protein in time and space

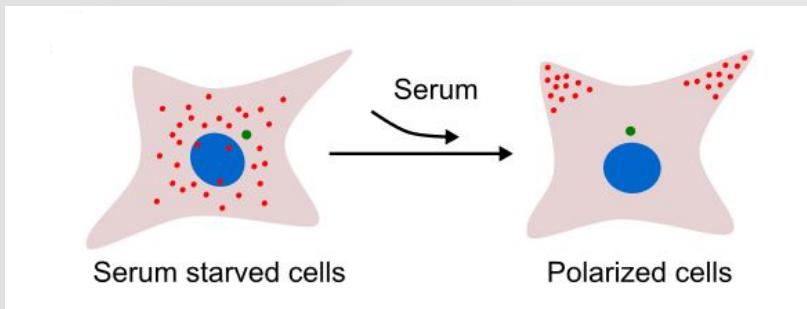
# Statistical analysis



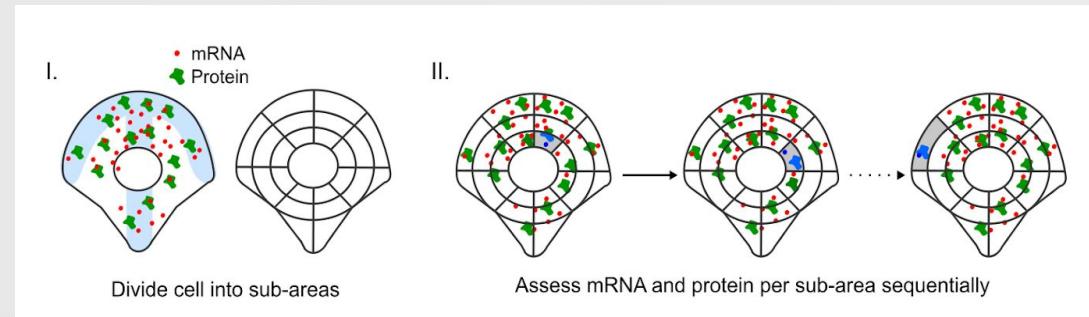
Series of images with mRNA position and protein signal intensity, cell and nucleus area, we can compute some statistics:

- Quantitative information on mRNAs and proteins
- Diffusion in the cytoplasm
- **Diffusion around the nucleus - at the cell periphery**
- Diffusion according to MTOC's position
- Clustering mRNA / Protein in time and space

# Peripheral enrichment



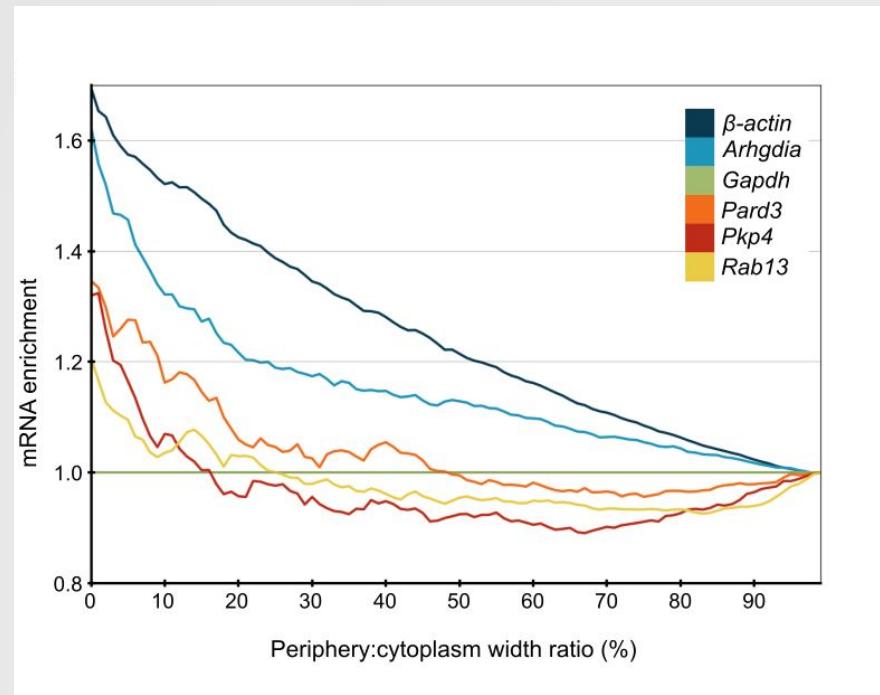
Peripheral enrichment and clustering dynamics of mRNA at the leading edge of cells (Mouse fibroblasts).



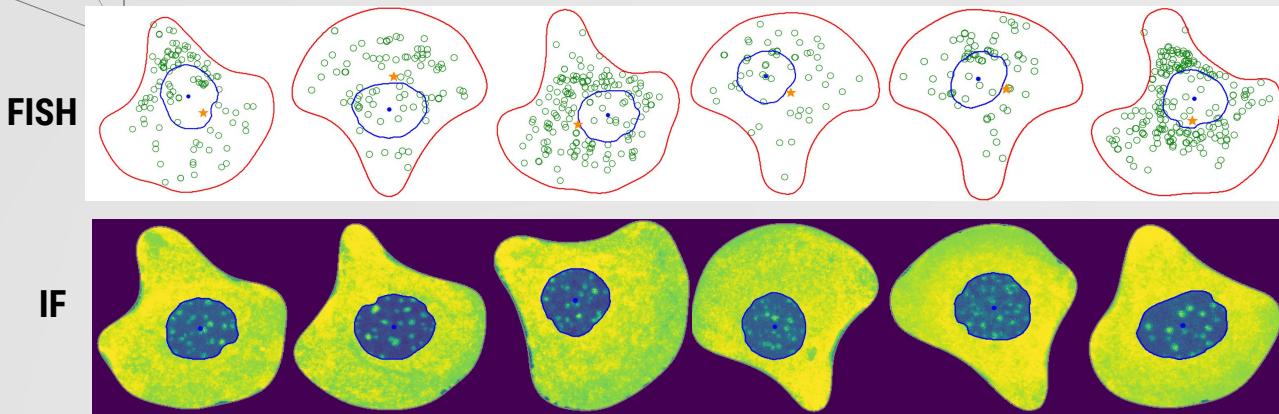
Fine grained quantization of cells.

# Peripheral enrichment

Comparison of the enrichment of 5 mRNAs with respect to Gapdh mRNA in a peripheral cellular region whose width varies from 0-100% of the radial distance from the plasma membrane to the nucleus.



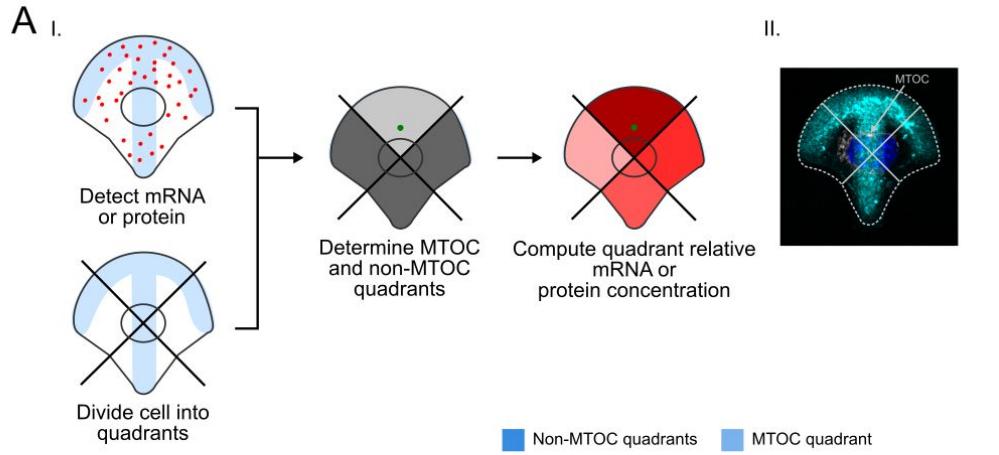
# Statistical analysis



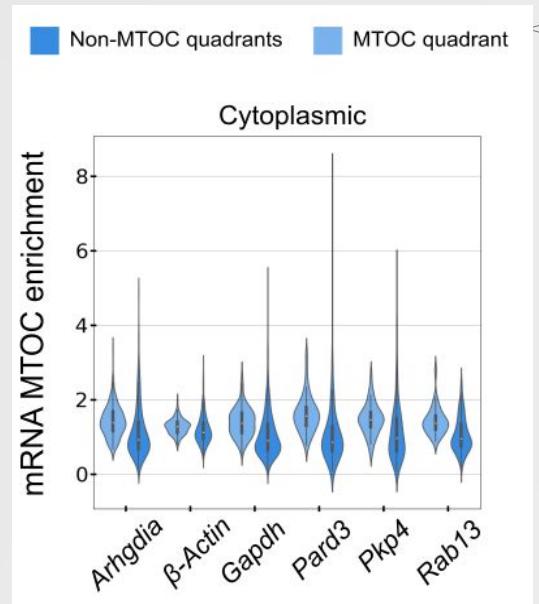
Series of images with mRNA position and protein signal intensity, cell and nucleus area, we can compute some statistics:

- Quantitative information on mRNAs and proteins
- Diffusion in the cytoplasm
- Diffusion around the nucleus - at the cell periphery
- **Diffusion according to MTOC's position**
- Clustering mRNA / Protein in time and space

# Colocalization

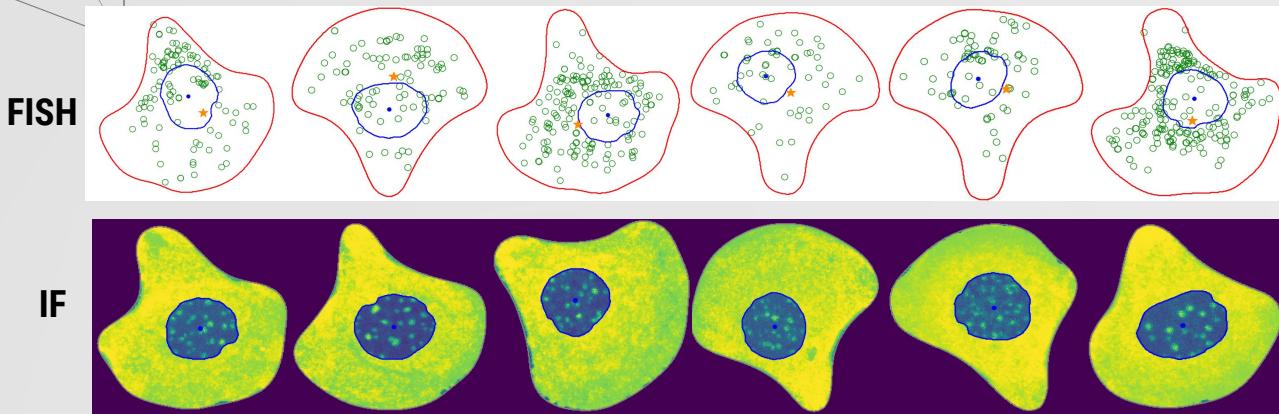


Divided cell into quadrants, used as regions over which we estimate mRNA and protein local relative densities



Higher enrichment of all cytoplasmic mRNA transcripts in the MTOC containing quadrant compared to the non MTOC-containing quadrant

# Statistical analysis



Series of images with mRNA position and protein signal intensity, cell and nucleus area, we can compute some statistics:

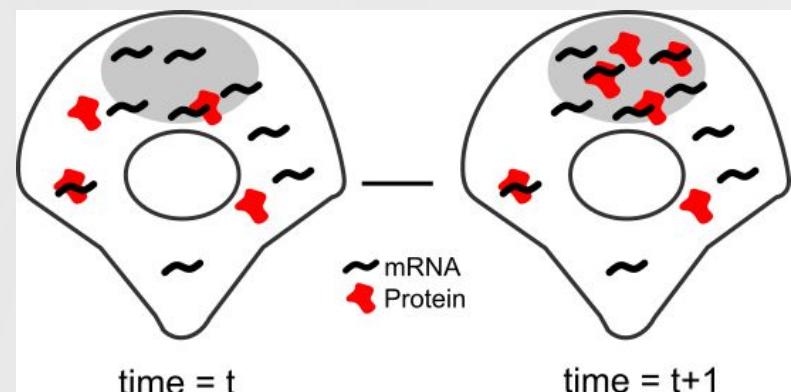
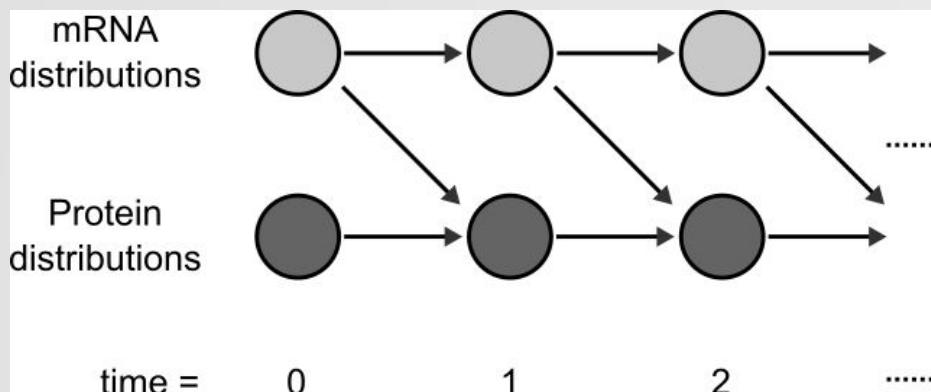
- Quantitative information on mRNAs and proteins
- Diffusion in the cytoplasm
- Diffusion around the nucleus - at the cell periphery
- Diffusion according to MTOC's position
- **Clustering mRNA / Protein in time and space**

# Statistical analysis

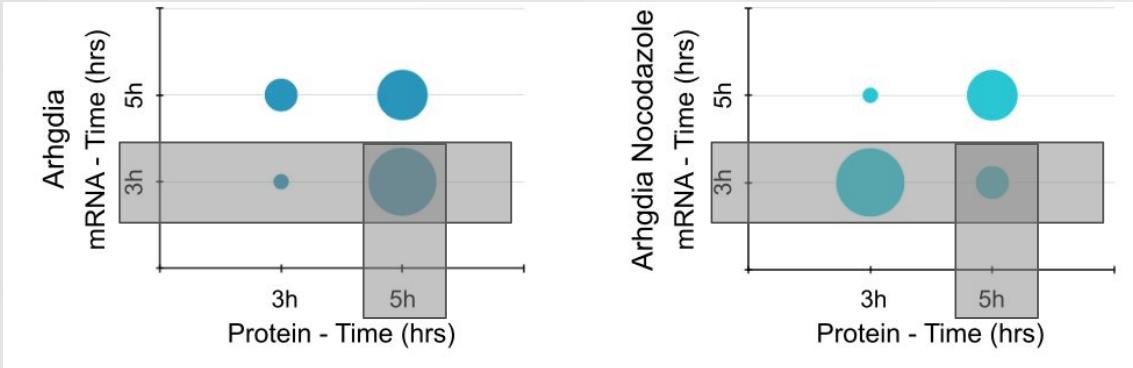
## Does mRNA position influence protein position over time?

Measure the interdependence between the mRNA and protein dynamic

Correlation between mRNA and protein spatial distributions at several time points



# Statistical analysis



Strong correlation  
Weak correlation

Nocodazole interferes with the polymerization of microtubules

Correlation between the subcellular localization of the mRNA and corresponding protein is altered with nocodazole

# Take home message

Deep Learning is possible even on "ugly" datasets (with effort)

Currently working to apply our model to other datasets

We are looking for collaborations to improve our models



## Data Acquisition

<http://dypfish.org/>



## Data Analysis

Soon on GitHub



## Statistical Analysis

<https://github.com/cbib/dypfish>

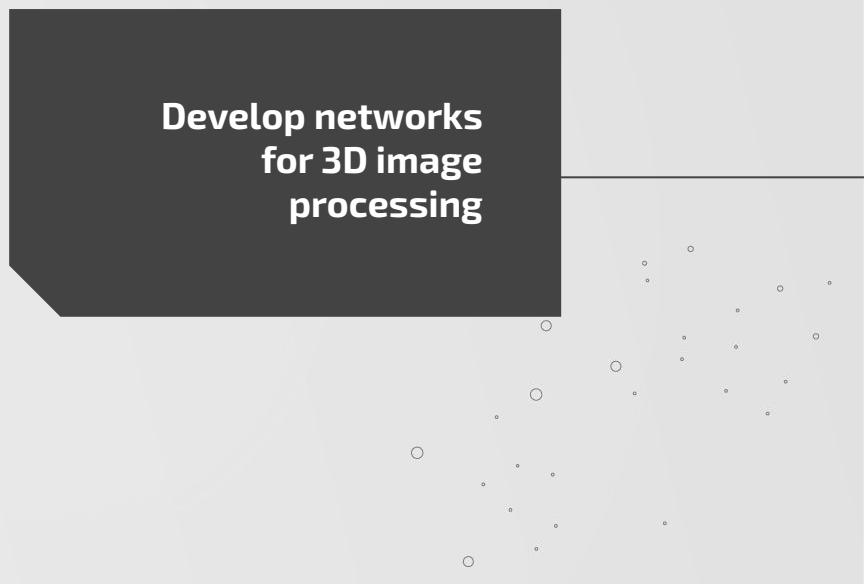




# Perspectives

**Upgrade our networks  
for spot detection  
(with less supervision)**

**Generalize our models  
for FISH signal  
detection**



**Develop networks  
for 3D image  
processing**



# cbib

BIOINFORMATIQUE



Macha Nikolski  
Head of CBiB



Benjamin Dartigues  
Bioinformatics Engineer



Edgar Lefevre  
R&D Engineer



Hayssam Soueidan  
R&D Engineer

[emmanuel.bouilhol@u-bordeaux.fr](mailto:emmanuel.bouilhol@u-bordeaux.fr)

# THANKS

mhlangalab

Musa Mhlanga  
Anca Savulescu  
Robyn Brackin  
Caron Jacobs

